

# Reinforcement Learning zur Ableitung autonomer Handelsstrategien in Elektrizitätsmärkten

Energiesystem- und Klimamodellierung  
Philipp REUBER<sup>1(1)</sup>, Jacob TRAN<sup>(1)</sup>, Simon KRAHL <sup>(1)</sup>, Albert MOSER<sup>(2)</sup>  
<sup>(1)</sup>FGH e.V., <sup>(2)</sup>RWTH Aachen University

## Motivation und zentrale Fragestellung

Die zunehmende Dargebotsabhängigkeit der Stromerzeugung sowie neuartige Verbraucher sorgen für eine steigende Volatilität des Strompreises an der Strombörse. Insbesondere die mittel- und kurzfristige Unsicherheit in der Erzeugung führt zu stark schwankenden Preisen, welche vor allem untertätig und zwischen zwei Handelstagen auftreten. Als Instrument zum Ausgleich dieser Abweichungen gewinnt der kurzfristige Stromhandel an Bedeutung. Die optimale Beschaffung oder Vermarktung von Strom stellt aufgrund vielfältiger Einflussfaktoren ein komplexes Optimierungsproblem dar und erfordert die Entwicklung geeigneter Handelsstrategien. Nachfolgend soll daher ein Verfahren basierend auf dem Reinforcement Learning vorgestellt werden, welches die autonome Ableitung von Handelsentscheidungen unter Berücksichtigung der Wechselwirkungen ermöglicht. [1-2]

## Methodische Vorgangsweise

Der elektrische Energiehandel erfolgt im Wesentlichen per „Over The Counter“ (OTC)-Handel oder über die Strombörse. Während an der Börse feste Handelszeiten existieren können OTC-Handelsgeschäfte jederzeit ausgeführt werden. Der Preis an der Börse wird durch die Kosten des letzten zur Deckung der Nachfrage herangezogenen Kraftwerkes bestimmt und steht somit erst am Ende der Auktion fest. Da bereits vor Auktionsende Geschäfte im OTC-Handel möglich sind, kann der Handelspreis dort von dem späteren Börsenpreis abweichen. Diese Preisdifferenzen können zur Gewinnerzielung verwendet werden, wenn ein Handelsgeschäft im OTC-Markt durchgeführt und ein gegenläufiges Handelsgeschäft an der Börse platziert wird. In dieser Arbeit wurde daher ein Agent mittels Reinforcement Learning in der beschriebenen Handlungsumgebung trainiert, um Preisdifferenzen zur Gewinnmaximierung auszunutzen. [1-3]

Reinforcement Learning versucht den natürlichen Lernprozess durch Ausprobieren nachzubilden. Hierbei soll der Handelsagent auf Grundlage von Eingabedaten der Umwelt Entscheidungen treffen. Das Modell bewertet den aktuellen Preis im OTC-Markt und trifft auf Basis der zuvor ausgewählten Eingangsdaten eine Handelsentscheidung. Nach Börsenschluss wird der realisierte Handel hinsichtlich des Gewinns oder Verlusts bewertet. Durch das Feedback werden die gewählten Handelsentscheidungen evaluiert und optimiert. Die Handelsstrategie des Agenten wird in diesem Verfahren durch künstliche Neuronale Netze (KNN) bestimmt. In einem iterativen Prozess erlernt ein solches KNN die Zusammenhänge zwischen den Eingabedaten und deren Auswirkungen auf den Strompreis.

## Ergebnisse und Schlussfolgerungen

Die Validierung des Verfahrens erfolgte durch exemplarische Untersuchung anhand von historischen Daten für die Deutsche Gebotszone im Umfang von eineinhalb Jahren. Von diesem Zeitraum werden 175 zufällig ausgewählte Tage zur Validierung des Verfahrens herangezogen. In Tabelle 1 sind die Gewinne für beide Datensätze für verschiedene Strategien dargestellt. In der Strategie „Immer kaufen“ bzw. „Immer verkaufen“ wird für jeden Handelstag immer ein Kaufgeschäft bzw. Verkaufsgeschäft im OTC-Markt platziert. Diese konstanten Strategien dienen der verbesserten Einordnung der Ergebnisse des Verfahrens, Zur besseren Vergleichbarkeit ist die Anzahl der Handelstage je Datensatz und der maximal erreichbare Gewinn angegeben.

Als Handelsprodukt werden Stromlieferungen für einen gesamten Tag (Base) verwendet. Zur Vereinfachung des Modells kann der Agent im OTC-Markt nur ein einziges Geschäft platzieren. Dieses wird dann mit einer gegenläufigen Transaktion mit dem späteren Auktionspreis ausgeglichen.

---

<sup>1</sup> Jungautor, Roermonder Str. 199, 52072 Aachen, +49 241 997857-148,  
[philipp.reuber@fgh-ma.de](mailto:philipp.reuber@fgh-ma.de), <https://fgh-ma.de/>

Somit wird versucht, die Preisdifferenz zwischen beiden Marktplätzen auszunutzen. Das Handelsvolumen wurde auf 1 MW festgelegt.

Tabelle 1: Vergleich der Gewinne für die jeweiligen Datensätze für konstante Strategien

	Trainingsdatensatz	Testdatensatz
Immer kaufen	-35,34€	22,86€
Immer verkaufen	11,34€	-28,57€
Maximal möglicher Gewinn	922,22€	248,44€
Anzahl Tage	702	175

Abbildung 1 zeigt, dass sowohl der Trainingsgewinn und der Testgewinn positive Werte annehmen. Ein Vergleich mit den Strategien aus Tabelle 1 zeigt, dass die konstanten Strategien deutlich hinter den durch das Verfahren generierten Gewinn zurückbleiben. Folglich ist dieses Verfahren zur Ableitung von Handelsstrategien in Elektrizitätsmärkten geeignet. Jedoch zeigt sich, dass die Anzahl der Trainingsepisoden einen signifikanten Einfluss auf den Testgewinn hat.

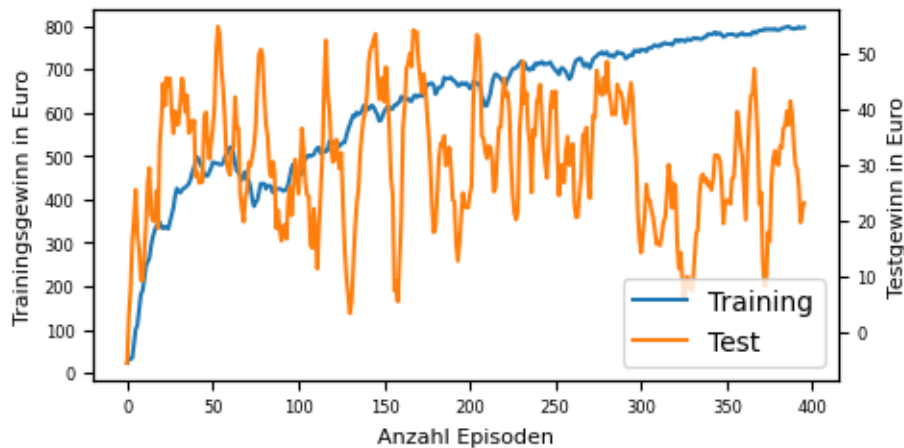


Abbildung 1: Verlauf des Gewinns für den Trainings- und Testdatensatz in Abhängigkeit der Episodenanzahl (Trainingsdurchläufe)

## Literatur

- [1] Schiffer, H.-W.: Energiemarkt Deutschland. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. ISBN: 978-3-658-23023-4. DOI: 10.1007/978-3-658-23024-1.
- [2] Schumacher, I.; Würfel, P.: Strategien zur Strombeschaffung in Unternehmen: Energieeinkauf optimieren, Kosten senken. Wiesbaden: Springer Gabler, 2015. ISBN: 978-3-658-07421-0. DOI: 10.1007/978-3-658-07422-7. URL: <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=976907>.
- [3] Konstantin, P.: Praxisbuch Energiewirtschaft. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. ISBN: 978-3-642-37264-3. DOI: 10.1007/978-3-642-37265-0.